

CS 44800: Introduction To Relational Database Systems

Prof. Chris Clifton
24 August 2021



Why this Course?

- Managing Data is one of the primary uses of computers
- This course covers the foundations of *organized* data management
 - Database Management Systems (DBMS): Tools to safely store and work with large quantities of information
- Much success in research
 - Relational theory spawned numerous products and companies
- But still lots to do
 - Can we get the other 90% (?) of data into databases?

Database Management Systems

- Query
 - Find the data by *properties*
 - Statistical summaries of data
- Consistency
 - Prevent errors
 - Consistency between objects
 - Consistency with queries/updates
- Reliability
 - Data isn't lost/damaged
- Performance
 - Rapid queries/updates

What is a Database?

- Collection of data, used to represent the information of interest to one or more applications in a given organization
 - Usually large
 - Organized for rapid search and retrieval
- Database Management System (DBMS):
Tool to ease construction of databases
 - (Vendor) definition of database: Collection of data managed by a DBMS
- Properties:
 - Persistent Storage
 - A File System does this*
 - Query Interface
 - Information retrieval system*
 - Transaction Management

DataBase Management System (DBMS)

A system (**typically software**) able to manage data collections that are:

- **Large:** the data sizes are typical much larger than the capacity of computer main memories; today, because of the presence of multimedia data, the database sizes can be huge
- **Persistent:** the data last for a (possibly very long) period of time which is independent from the executions of the application programs that create and use the data
- **Shared:** used by different applications and users

DBMS

- A DBMS must assure:
 - **Reliability:** the data must survive to hardware and software errors
 - **Confidentiality:** access to data must be controlled
- As the majority of computer systems, a DBMS must be **efficient** (by optimizing the resources of the underlying system) and **effective** (by allowing users to make productive use of data)

Motivations for DB Technology

Organization/Enterprise

- It uses a set of resources, policies and regulation to execute the activities of interest for its own goals
 - Information and knowledge represent a key resource
- The *information system* is today always present in any organization we may think of
- The information system executes/manages information processes, that is, processes that involve information

Motivations for DB Technology

Information Systems

- It is a component (subsystem) of a given organization.
- It manages (that is, acquires, processes, stores, produces, and delivers) information of interest to the organization.
- Each organization has an information system, even though such system may not be explicitly present the organization structure
- In most cases, the information system supports other subsystems in the organization and therefore it may have to support users and applications from different sectors of the organization
- The information system is usually organized in subsystems (with a distributed or hierarchical organization); these subsystems may be tightly or loosely coupled

Motivations for DB Technology

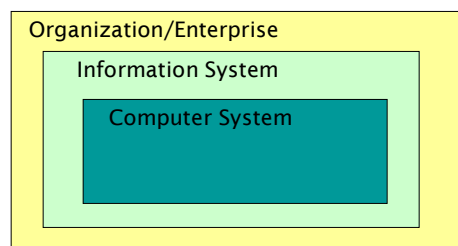
Automation of Information Systems

- The concept of “information system” is independent from IT technology: there are organizations, the goal of which is to manage information (like in the case of demographic services), and that have been in place for centuries

Motivations for DB Technology

Computer system

- Automated portion of the information system
- Component of the information system that manages information through the use of computer systems



Motivations for DB Technology

Database systems

- Database systems represent the most important computer technology for implementing and supporting information systems
- Therefore, they represent a key technology for any modern organization/enterprise
- Because information are a key resources, database systems must be reliable, secure and efficient

Motivations for DB Technology

Information and Data

- Computer systems represent information through data
- *Data* represents information. *Information* is the (subjective) interpretation of data
- **Data** - Physical phenomena chosen by convention to represent certain aspects of our conceptual and real world. The meaning we assign to data are called information. Data is used to transmit and store information and to derive new information by manipulating the data according to formal rules.

from:

P.Brinch Hansen. *Operating Systems Principles*.
Prentice-Hall, 1973.

Motivations for DB Technology

Information and Data

- The data are elementary facts that need to be interpreted in order to convey information
- Example

Consider a data item represented by the integer number 3; such number does not provide any information

By contrast, saying that 3 is the number of credits of CS541 provides some information

Motivations for DB Technology

Information and Data

- One of the fundamental goals of a database management system (DBMS) is to provide an interpretation context to a collection of data, so that users can effectively access information encoded by this collection of data

Relational Database Management Systems

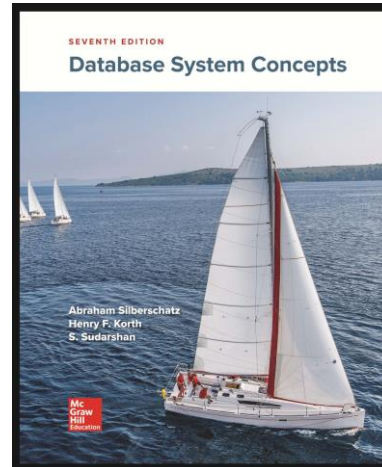
- Mathematical data model
 - Used properly, ensures some forms of consistency
- Mathematical (set based) query language
 - Declarative
 - Query optimization/processing
- Many alternatives have been proposed
 - Relational has incorporated their key benefits

Logistics

- Time and location: TR 13:30-14:45, KRAN G016
 - Live remote attendance via Zoom (link found in Brightspace)
 - Lectures recorded and available in Mediaspace (Look in Brightspace, "Kaltura Media Gallery")
- Instructor: Chris Clifton, clifton@cs.purdue.edu
HAAS 222, office hours: TBD
- Teaching assistants:
 - Iram Tazim Hoque ihoque@purdue.edu
 - Abdullah Al Mamun mamuna@purdue.edu
 - Faisal Tariq Vora voraf@purdue.edu
 - Libin Zhou zhou822@purdue.eduOffice hours: times TBD. Some will be available on WebEx or Zoom
- Webpage: <http://www.cs.purdue.edu/~clifton/cs44800/>
- Email list: fall-2021-cs-44800-LE1@lists.purdue.edu
 - Goes to your @purdue.edu email, make sure you get this, as it contains critical announcements.
- Discussion forum: Piazza? Brightspace? Others?
- Prerequisites: CS25100. Prior experience with database (e.g., knowledge of SQL) helps

Readings

- *Database System Concepts, Seventh Edition*
Avi Silberschatz, Henry F. Korth, and S. Sudarshan
McGraw-Hill, ISBN 9780078022159, 2019



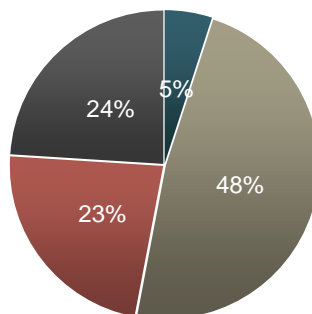
PSOs

- PSOs are used to review and go deeper into material
 - Prepared material, not just a Q&A session
 - *Smaller size means it is easier to discuss than in lecture*
- Will often recap lecture material in ways tied directly to projects / assignments
- May also be an opportunity to collaboratively work with your peers

Workload

- **Homeworks**
 - Mix of written assignments and programming projects
Over the course of the semester you will build much of an RDBMS
 - Late policy: 15% off per day late, maximum of 5 days
 - Five extension days to be used at your discretion
 - No fractional days
 - May not be used to extend submission past last day of class.
 - Late penalties will not be applied until the end of the semester
 - Late days will be applied to your best advantage
- **Exams**
 - Midterms (2, in class) and final exam, in person
 - *Arrangements made for quarantine/etc. on a case-by-case basis*

Grading: *Tentative, will be adjusted depending on how the semester goes*



■ Participation ■ Homework ■ Midterms ■ Final

Classroom Interaction Preferences

- Uses
 - Feedback to instructor: Do you understand what I'm trying to present, or do I need to go over it again?
 - Participation: Did you respond?
 - Knowledge and Understanding: Did you get it right?
Only a few questions will be scored, in many cases I may ask questions to get you thinking and don't expect you to know the answer
- Methods
 - and how to do it in a manner that is fair to people who may be quarantined*
 - [iClicker](#)
 - [HotSeat](#)
 - [TopHat](#)
 - *Others?*

21

Recommendations

- Form study groups, collaborate with your peers
 - *Some projects may be done in groups*
 - *Groups may be assigned, not your choice – just as in the working world*
- Collaboration: Goal is to be able to do the work yourself
 - Discuss how to solve problems
 - But work out the answer yourself
- Unless explicitly a group task, what you turn in should reflect your own work, and ability to do similar problems on your own

22